



13th COTA International Conference of Transportation Professionals (CICTP 2013)

Analyzing the Activity Areas of Non-resident Tourists of Shanghai Expo using Cellular Phone Data

Fengjiao Jia^a, Xiaoyun Cheng^a, Zhengyu Duan^{a,*}^a*Key Laboratory of Road and Traffic Engineering of the Ministry of Education, Tongji University
4800 Caoan Highway, 201804, Shanghai, China*

Abstract

With the continuous development of Metropolitan Coordinating Region, the core cities in the region tend to have some famous tourism resources, such as the 2010 Shanghai Expo. As the most important destinations in the city, these resources will generate much relevant traffic demand, which will deteriorate the traffic condition of urban interior transport system. For example, the non-resident tourists may visit Nanjing Road East or Yu Garden after they leaving the Expo site, increasing the traffic demand in those areas. Mobile Phone data offer us a chance to study the behaviour of a tourist all day long, which will help us to improve the existing facilities and service and adjust the traffic resources according to the real time transportation demand. This paper uses the statistical method to find the main activity areas of non-resident tourists and the spatial distribution of crowd in Shanghai during day on the Mobile Phone. We come up with a methodology to identify the Non-resident Tourists in Shanghai Expo and calculate the residence time of the base station. The goal of this paper is to analyze the scale and scope of this kind of relevant traffic demand and provides a basis method for monitoring tourists' behaviour impacts using mobile phone-based data in the future.

© 2013 The Authors. Published by Elsevier Ltd. Open access under [CC BY-NC-ND license](http://creativecommons.org/licenses/by-nc-nd/4.0/).

Selection and peer-review under responsibility of Chinese Overseas Transportation Association (COTA).

Keywords: Activity Area; Non-resident Tourists; Mobile Phone Data; Expo

1. Introduction

With the continuous development of Metropolitan Coordinating Region, the core cities in the region (such as Shanghai in China's Yangtze River Delta) tend to have some famous tourism resources. For example, Shanghai has successfully hosted the 2010 Expo, and will have another famous tour resource—Disneyland. Many travellers

* Corresponding author. Tel.: 86-21-69583775;
E-mail address: d_zy@163.com

choose to visit these kinds of destinations, and then go to some other tourism attractions. In other words, these core resources will generate much relevant traffic demand, which will deteriorate the traffic condition of urban interior transport system. We are facing the challenge to analyze the scale and scope of this kind of relevant traffic demand.

Method for analyzing tourists' behaviour impacts using mobile phone-based data is really different from other traditional survey methods, such as questionnaire. Questionnaire survey cost too much money and time and we can't get the results instantly. The attractive advantage of the data from a mobile phone network is that it is massive and cheap (we just need to establish a database to collect the data and don't need more people to distribute the questionnaires and input data); the tourists' behaviour can be updated in real time. Many researchers have used mobile phone-based data to analyze the travel behaviours.

From the point of view of the precision, Cell phone positioning includes two kinds of methods. One is obscure method (precision: 50~300m), such as Cell-ID, TDOA, TA and Handover. The other is accurate method (precision: 5~30m), for instance, E-OTD, U-TDOA, A-GPS and GPS One.

Many researchers have proposed some methods to determine the person's path. F Calabrese et al. (2010) defined a trajectory as a sequence of chronological location points for each user and found where they stopped travelling by determine a spatial threshold and a certain time interval[1]. Pu Wang et al. (2009) broke each day into 24 hour long time intervals and chose the location with a probability that in proportional to the frequency of the towers observed in that time period[2]. Katayoun Farrahi and Daniel Gatica-Perez (2008) discovered people's routines by setting up a Topic Framework [3]. Hao Tian et al. (2010) proposed LSI and Location stability concept, in order to estimate a person's space-time path.[4]

Mobile Phone data can also be used to obtain the information about Urban Land use. Li Xiaopeng et al. (2005) introduced a new method to analyze and classify urban activities and traffic properties of urban lands [5]. Carlo Ratti et al. (2006) took Milan as an example, and their results enabled a graphic representation of the intensity of urban activities and their evolution through space and time [6]. Zhengyu Duan et al. (2011) analyzed the Land use of Shanghai and generated Origin-Destination matrices from millions of individual mobile phone traces [7].

Within the topic of the analysis of people's behaviour, Marta C. Gonzalez (2008) found that human trajectories show a high degree of temporal and spatial regularity [8]. Chao ming Song et al. (2010) discovered that although the distances of the trips were different, 93% of the behaviour could be predicted and remain the same [9]. Sibren Isaacman et al. studied human mobility in Los Angeles and New York, and included that human mobility changes with the seasons [10]. Shang Wang (2011) refined and analyzed the service radius of the Urban Railstation[11]. Andres Sevtsuk et al. (2010) determined the aggregate distribution of Rome's population over time by using the volume of call activity in mobile network cells as the unit of spatial analysis [12]. Chaogui Kang et al. (2010) analyzed 9-day person's behaviour, and introduced several pre-processing and spatiotemporal analysis methods for new ICT data in individual human mobility patterns mining and urban analysis[13]. Santi Phithakkitnukoon (2010) found a strong correlation in daily activity patterns within the group of people who shared a common work area's profile[14].

Non-resident tourists are the people who live in other cities and come to visit the core city. Compared with the local people, they have some special traits. They stayed in the city for just a few days, and visited many places during a short time. The usual vehicles used by the tourists are mass transit, such as bus, tube and taxi. In this paper, we use the statistical method to find the main activity areas of non-resident tourists outside the Expo Site and the crowd distribution during different parts of a day in Shanghai based on the Mobile Phone data. We try to analyze the scale and scope of relevant traffic demand and provide a basis method for analyzing tourists' behaviour impacts using mobile phone-based data, which can be used to analyze the traffic demand in the future.

2. Description

The datasets used in our paper consist of anonymous mobile phone signalling data collected by one of the wireless operators(70% of market share), and it was collected from more than twenty thousand base stations (BS),including about 1.4 billion operation records every day(24 hours per day). The collecting time was 5 days (July 11st to July 5th), 2010, during the holding time of Expo.

Every record includes five parts:

1. MSID: Unique identity number for a Mobile user
2. TIMESTAMP: Timestamp of the record, units: ms
3. LAC: Number of Location area
4. CELLID: Number of the base station
- 5.EVENTID: 11 Types of signaling, including Random location update, Periodic location update procedure, CM Service Request, Paging Response, BSC Handover and so on.

Table 1 provides some examples of the data.

Table 1.Attribute of the Base Stations (Example)

LAC	CELLID	Longitude	Latitude
6289	45457	121.046	30.895
6358	1041	121.693	31.170
6295	145	121.437	30.895
6299	53889	121.464	31.231
6262	41489	121.429	31.156

According to the locations of the BSs, we divided the BSs into two different groups .Some BSs, which were called Expo-BSs, were in the area of the Expo. Others were called Non-Expo BSs, because they were found outside the Expo area.

3.Research sample

3.1.Screening criteria for Research Sample

The research objects are the non-resident visitors who had been to the Expo. The first task is to select the non-resident visitors, and the second work is to find who had visited the Expo among those non-resident visitors.

A survey said the short-time travel time of a visitor who visits Shanghai is shorter than 3 days, so we find the people whose mobile phone signalling data appeared on several successive days(less than 3 days) as the non-resident visitors. In order to improve the accuracy of classifying, we ignore the tourists who had appeared either in the first day or the last day because we cannot tell the accurate time when they came and left Shanghai.

In order to screen out the tourists who had visited the Expo, we need to establish a set of criteria. In our survey, we get the research objects using the following way:

1. The effective visiting time of Shanghai Expo is between 9:00AM and 12:00PM.
2. For every non-resident visitor, after searching the maximum time (t_{\max})and the minimum time(t_{\min}) of the Expo BSs signalling data (during a day), we calculated $\Delta t = t_{\max} - t_{\min}$, and analysed the BSs appeared between t_{\max} and t_{\min} . There are two circumstances (Figure 1):

- 1) None of BSs during this period is the Non-Expo Base station. Then

$$\Delta t = t_{\max} - \max(t_{\min}, 9:00) \quad (1)$$

If $5h \leq \Delta t \leq 12h$ then, this visitor is the research object.

2) There are some BSs are the Non-Expo BSs. Then

$$\Delta t' = \Delta t_1 + \Delta t_2 + \dots + \Delta t_n, \quad \Delta t_i = t_{\max i} - \max(t_{\min i}, 9:00) \quad (2)$$

If $\Delta t' \geq 0.8 \times \Delta t, 5h \leq \Delta t \leq 12h$, this visitor is the research object.

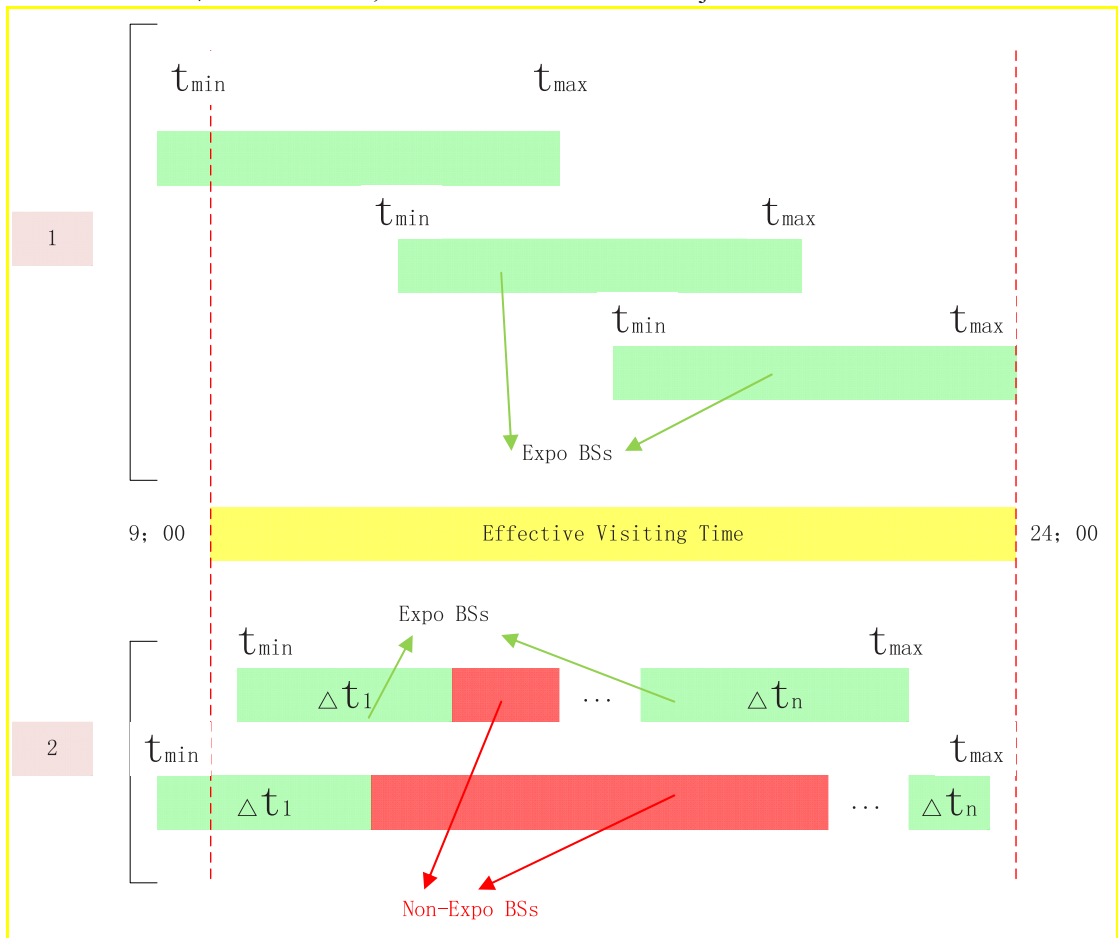


Fig. 1. Schematic Diagram of Screening Out Expo Tourists.

3.2. Research Sample of Non-resident Tourists

Then, we find our sample size is 139978.

4. Methodology

4.1. Identify The Trip Chains of Non-resident Tourists

Because of the complexity in analyzing the tourists' behaviour, our research is based on the following hypothesis:

1. Mobile phone users' locations of mobile records overlap with the base station location, and on behalf of the users' real station.

There are many positioning methods based on mobile phone, such as CELL-ID, TDOA (Time Difference of Arrival), Handover and E-OTD (Enhanced Observed Time Difference), and each method has its own positional Accuracy. Scale and scope of relevant traffic demand is a macroscopically problem, so we don't need to know the exact position of every tourist. CELL-ID has already qualified for the position precision, so we choose this method to position tourists.

2. Sampling Interval has no effect on the analysis of personal behaviour.

Sampling Interval of every tourist's record is not uniform distribution, which means that the numbers of operation records collected from each tourist are not the same. According to Albert-Laszlo Barabasi research, Sampling Interval of Mobile data has little influence on the research of the distance and mode of human behaviour. [8]

We extracted the trip chain of every tourist linking with the coordinate of every base station:

$$R^a = \{\overline{r_1^a}, \overline{r_2^a}, \overline{r_3^a}, \dots, \overline{r_n^a}\}, \overline{r_i^a} = (x_i, y_i, t_i) \quad (3)$$

Where R^a is the trip chain of a^{th} tourist, $\overline{r_i^a}$ is the attribute of i^{th} data, (x_i, y_i, t_i) the coordinate and timestamp of the i^{th} data.

After extracting and drawing the trip chains of the tourists, we find a significant character of the trip chains of the tourists:

The residence times of some places were longer than other areas. That means those places influent the tourists' trip differently and we can find the main activity areas easily. Then, we should calculate the residence as the weights of every Base Stations.

4.2. Compute the Residence Time of Non-resident Tourist

The data in our research include 11 types of signalling, which help us to calculate the residence time of each base station. If the BS of the record differs from the BS of the last record, we can compute the residence time of the last record, using the time interval between these two records. Table 2 is an example to explain how to calculate the Residence Time of a BS.

Table 2. Computing the residence time of a BS

ID	MSID	TIMESTAMP	LAC	CELLID	Residence Time(s)
1	000841EB45DA883	2010,0715,21:02:43	6164	18690	—
2	000841EB45DA883	2010,0715,21:02:43	6164	18690	215
3	000841EB45DA883	2010,0715,21:06:18	6163	4673	615
4	000841EB45DA883	2010,0715,21:16:33	6183	16450	17
5	000841EB45DA883	2010,0715,21:16:50	6163	1553	663
6	000841EB45DA883	2010,0715,21:27:53	6183	16770	2295
7	000841EB45DA883	2010,0715,22:06:08	6167	29443	—

TIMESTAMP: 2010, 0715, 21:02:43="21:02:43, July 15th, 2010"

—: BS of this record is same with that of the last record; do not need to be reserved

4.3. Get the Activity Area of Non-resident Tourists

Our purpose is to find out the main activity areas of non-resident tourists and the crowd distribution in Shanghai during different parts of the day. Some BSs are on the way to the destinations, meaning that the tourists did not stay around those BSs; they just went through those areas (through-BS), and some BSs are in the main tour areas of the tourists (stay-BS). We need to determine the interval T to tell which BS is not a through-BS. If the residence time of the BS is longer than T , we call it a stay-BS; if the residence time of the BS is shorter than T , we call it a through-BS. Finally, we use the stay-BSs as the activity areas in our research. In our study, we determine $T=15\text{min}$. Table 3 is an example to explain how to identify the Stay-BS.

Table 3. Identifying Method of Stay-BS

MSID	TIMESTAMP	LAC	CELLID	Residence Time(s)	Category of BS
000841EB45DA883	20100715210243	6164	18690	215	through-BS
000841EB45DA883	20100715210618	6163	4673	615	through-BS
000841EB45DA883	20100715211633	6183	16450	17	through-BS
000841EB45DA883	20100715211650	6163	1553	663	through-BS
000841EB45DA883	20100715212753	6183	16770	2295	stay-BS
000841EB45DA883	20100715220608	6167	29443	0	through-BS
00087F5E63CF0B2	20100714074614	6289	46371	314	through-BS
00087F5E63CF0B2	20100714075128	6250	5122	597	through-BS
00087F5E63CF0B2	20100714080125	6148	8833	205	through-BS
00087F5E63CF0B2	20100714080450	6152	12675	172	through-BS
00087F5E63CF0B2	20100714080742	6145	8451	507	through-BS
00087F5E63CF0B2	20100714081609	6252	16483	0	through-BS

5. Analysis of activity characteristic

After we pick out the records belong to a non-resident tourist who has visited Shanghai Expo, we find the main activity area of different groups.

5.1. Main Activity Area

We draw the stay-BSs in the following pictures (Figure 2).

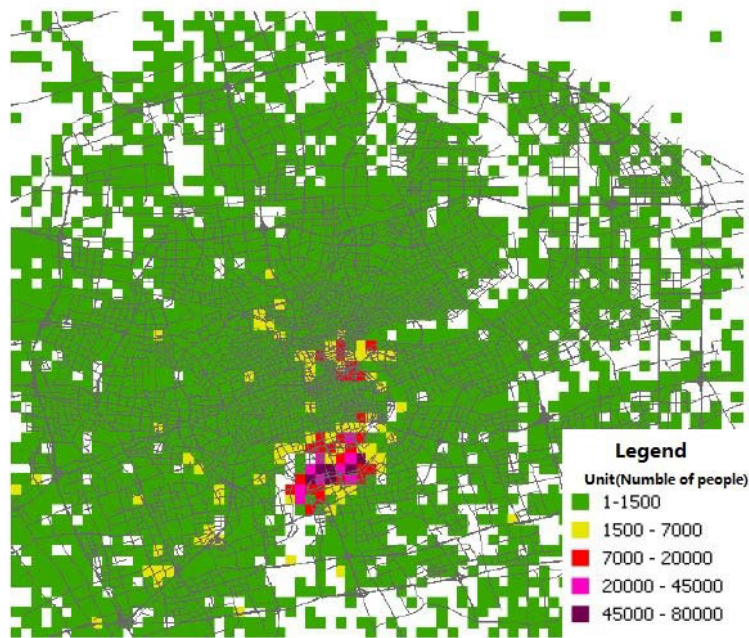


Fig. 2. Main Activity Areas of Different Groups.

We find that the main activity areas of the tourists are the Expo, The Bund, East Nanjing Road and some other famous scenic spots.

5.2. Activity Area during different time of a day

The non-resident tourists would visit different places during the day, so the distribution of the tourists may affect the transportation condition in the city. In order to avoid the traffic problems and provide enough traffic resources, we need to know where and when they visited during a day. Because the samples are the tourists who did not appear on the first day or the last day, we just draw the activity areas of the non-resident tourists at 4 different time of the day: 8a.m., 12a.m., 4p.m., and 8p.m. on July 12nd, 13rd and 14th (three days between the first day and the last day). Then, we choose to depict the activity areas on July 12nd as an example (Figure 3).

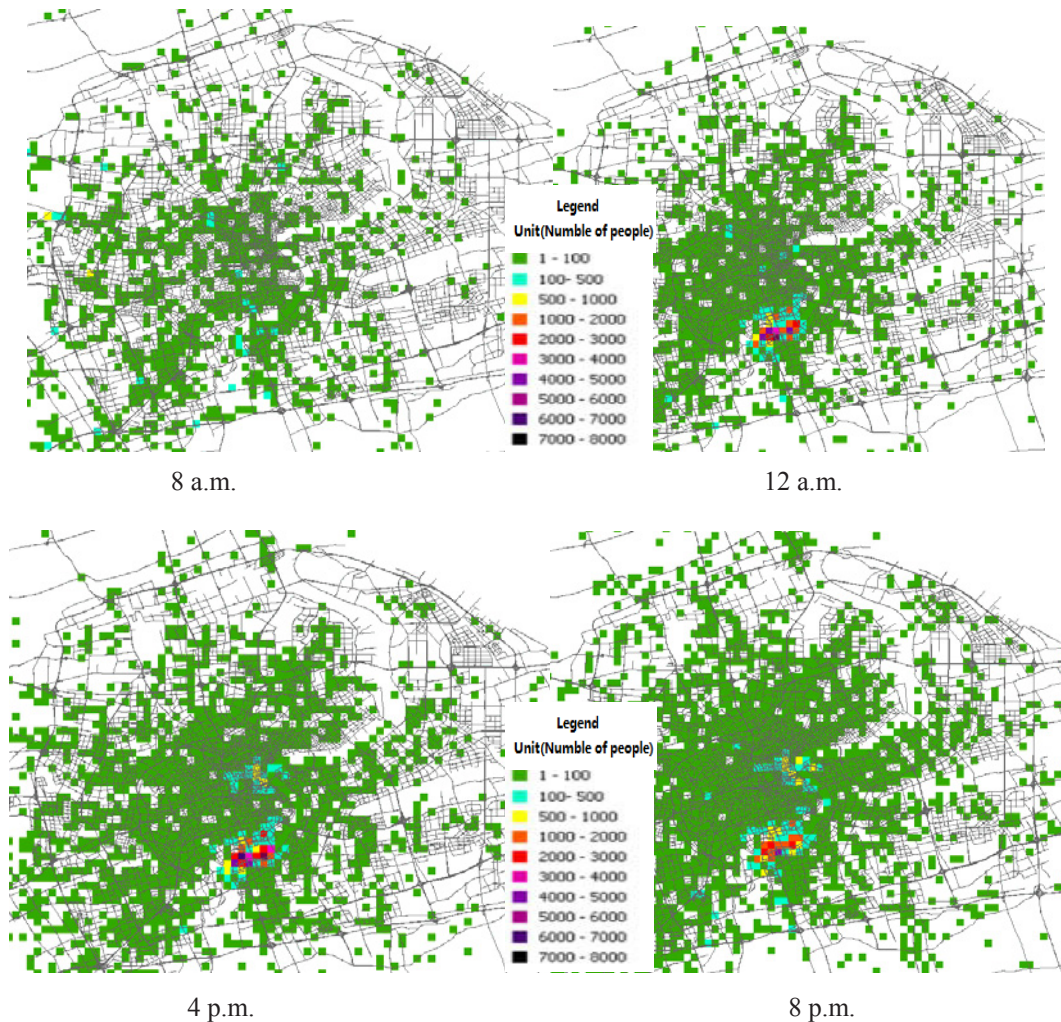


Fig. 3. Activity Areas of Non-resident tourists on July 12nd

The charts above show a significant trend of the main activity areas during the day. Obviously, the main area visited by the non-resident tourists was Expo. In the evening, more people chose to visit other places in the City, East Nanjing Road, Oriental Pearl TV Tower, etc.

From the analysis above, many tourists chose to visit East Nanjing Road and Yu Garden Area for shopping and other entertainments. But During the Expo Time, most of the taxis were arranged to evacuate the tourists around the Expo Area at night, making it was difficult to call a taxi around the Bund Area. If we can monitor the activity area of the tourists by the mobile phone data, we may provide more transportation resources to those areas effectively and efficiently, in order to keep the normal traffic pace in the city.

6. Conclusion

For a core city of Metropolitan Coordinating Region, such as Shanghai, the behaviour of the non-resident tourists has more important effect on the transportation system than years ago. Some famous resources attract thousands of tourists, and they may generate much relevant traffic demand, which will deteriorate the traffic condition of urban interior transport system.

As a new method, mobile phone-based data began to be used in the traffic analysis. The aim of our research is to investigate the non-resident visitors' behaviour during the period of Shanghai Expo. By this means, we can find the main activity areas during different parts of a day of the non-resident tourists, helping the transportation administrations to adjust the traffic resources to solve the traffic problems which were generated by the non-resident tourists.

Based on our analysis of the behaviour of the non-resident tourists whose had come to visit the Expo, we observed the following results:

Most of the non-resident tourists who had visited the Expo preferred to go to the Expo in the morning and stay there until late afternoon. When they left the Expo, some of them liked to visit other tourism places, such as East Nanjing Road, Yu Garden and Oriental Pearl TV Tower.

The collecting time of the mobile phone data is only 5 days. Although the non-resident tourists usually stay in Shanghai less than 3 days, we cannot avoid some deviation in the results. In future research, we need more data to prove the validity of the results.

Mobile phone-based data offer us a chance to trace the trip chain of a person easier. In the further, we can use them to investigate more results in other researches. For example, Shanghai Disneyland, which can generate much relevant traffic demand, will open in a few years. By monitoring the activity area of the tourists on the mobile phone data, the traffic management department can dynamically adjust the transportation resources, meeting the demand of the tourists.

Acknowledges

This work is supported by the Fundamental Research Funds for the Central Universities (Study of Urban Mobility Based on Mobile Phone Data).

References

- [1] F Calabrese, F Pereira, G Di Lorenzo et al. The geography of taste: analyzing cell-phone mobility and social events. *Lecture Notes in Computer Science*, 2010, Volume 6030/2010, pp.22-37
- [2] Pu Wang, Marta C. Gonzalez, Cesar A. Hidalgo et al. Understanding the Spreading Patterns of Mobile Phone Viruses. *Science Express*. 2009. Published 2 April 2009
- [3] Katayoun Farrahi, Daniel Gatica-Perez. Discovering Human Routines from Cell Phone Data with Topic Models. *IEEE*. pp.29-32
- [4] Hao Tian, Xiujun Ma, Han Wang et al. A Novel Approach to Estimate Human Space-time Path Based on Mobile Phone Call Records. *IEEE. Geoinformatics*, 2010 18th International Conference, pp.1-6
- [5] LI Xiaopeng, YANG Xinmiao, CAO Jinxin. A Method for Analyzing Traffic Properties of Urban Lands Based on Mobile Phone Traffic. *Chinese intelligent transportation annual meeting 1st, 2005*, pp.1010-1016
- [6] Carlo Ratti, Riccardo M. Pulselli, Sarah Williams et al. Mobile Landscapes: Using Location Data from Cell Phones for Urban Analysis. *Environment and Planning B: Planning and Design* 33(5) pp.727-748
- [7] Zhengyu Duan, Liang Liu, Shang Wang. MobilePulse: Dynamic Profiling of Land Use Pattern and OD Matrix Estimation from 10 Million Individual Cell Phone Records in Shanghai. *Geoinformatics*, 2011 19th International Conference on
- [8] Marta C. Gonzalez, Cesar A. Hidalgo & Albert-Laszlo Barabasi. Understanding individual human mobility patterns. *Nature*. Vol 453, 5 June 2008
- [9] Chaoming Song, Zehui Qu, Nicholas Blummet et al. Limits of Predictability in Human Mobility. 19 February 2010, *Science*, 327, 1018 (2010)
- [10] Sibren Isaacman, Richard Becker, Ramon Caceres et al. Ranges of Human Mobility in Los Angeles and New York. 8th IEEE

International Workshop on Managing Ubiquitous Communications and Services.

- [11] Shang Wang. Identifying the Services Region of Urban Rail Transit Station Using Cellular Probe Data. Master dissertation. Tongji University
- [12] Andres Sevtsuk, Carlo Ratti. Does Urban Mobility Have a Daily Routine? Learning from Aggregate Data Mobile Networks. *Journal of Urban Technology*, 17:1, pp.47-61
- [13] Chaogui Kang, Song Gao, Xing Lin et al. Analyzing and Geo-visualizing Individual Human Mobility Patterns Using Mobile Call Record. *Geoinformatics*, 2010 18th International Conference, pp.1-7
- [14] Santi Phithakkitnukoon, Teerayut Horanont, Giusy Di Lorenzo. Activity-Aware Map: Identifying human daily activity pattern using mobile phone data. *Lecture Notes in Computer Science*, 2010, Volume 6219/2010, pp.14-25